

# Effective Risk Assessment in Resilient Communication Networks

Krzysztof Rusek<sup>1</sup> · Piotr Guzik<sup>1</sup> · Piotr Cholda<sup>1</sup> 

Received: 4 February 2015 / Accepted: 28 March 2016 / Published online: 12 April 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** The paper discusses business impact analysis in the context of resilient communication networks. It is based on the total (aggregated) penalty that may be paid by an operator when the services (identified with transport demands) provided are interrupted due to network failures. The level of penalty is expressed as a commonly accepted business risk measure, Value-at-Risk (*VaR*). First, the main concern over *VaR*, namely the theoretical lack of subadditivity, is discussed. The study shows that, in practice, disadvantages do not appear in resilient network design, and *VaR* can be used without the need to apply more complex and less informative measures. Second, a method for calculating the upper bound of the total penalty is presented. The assessment is performed for unprotected and protected services with a broad variety of compensation policies used to translate technical loss to monetarily expressed penalty. The proposed bounds are experimentally shown to be effective in comparison with alternative calculation methods, and also in the case when some of the assumptions taken during the modelling stage are not met.

**Keywords** Availability · Compensation policy · Continuity · Design · Downtime · Management · Reliability · Value-at-Risk

---

✉ Piotr Cholda  
piotr.cholda@agh.edu.pl

Krzysztof Rusek  
krusek@agh.edu.pl

Piotr Guzik  
guzik@kt.agh.edu.pl

<sup>1</sup> Department of Telecommunications, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland

## 1 Introduction

Random failures, such as link cuts or hardware faults, are destructive to networks, both from a technical and business viewpoint. Penalties can be imposed on operators due to breach of Service Level Agreements (SLAs). To counteract these problems, risk management is defined as “coordinated activities to direct and control an organization in response to risk” [1].

With the recent ‘beyond connectivity’ trend [2]—where the main concern is to focus on the entire communication service—interest in network risk management has increased. In order to deal with failures from the risk management viewpoint, it is necessary to address the parameters such as: *probability* of adverse events, and the *loss (impact)* incurred by them. The latter is here related to penalties paid to clients affected by failures.

If risk is dealt with using the methodological approach, it is exercised within the cyclic risk management framework [3]. The simplified structure of this kind consists of the following steps: (1) risk assessment, (2) planning the risk response, (3) response deployment, and (4) risk monitoring. *Risk assessment* consists of: (a) risk analysis, identifying failure scenarios, and (b) risk evaluation determining their probability and impact on business goals. Here we use probabilistic risk assessment, during which both parameters are expressed mathematically [4].

Although in this paper we focus on risk evaluation alone, we also outline shortly other phases of the cycle. Designers of resilient networks are typically most familiar with the *risk response* stage, since the task of the technician is to prepare response strategies. The manager of the service provider then decides which one to choose. In resilient networks, the basic approach is risk mitigation that involves decrease of the impact. In practice, mitigation uses combination of resilience procedures [5]. In this case, *response deployment* embraces the configuration of resources and testing. The next step, *risk monitoring*, includes: (a) continuous risk monitoring, where risks are observed in order to identify new ones, and (b) response monitoring, performed to check if the implemented response meets the intended goals.

A current practice of dealing with failures in design and management of resilient networks is based on measuring failure risk with purely technical methods. These methods apply steady-state availability or mean downtime as risk measures. However, they are not relevant in a business context. Firstly, it is more important to express the consequence of failures in monetary terms. Secondly, businesses may be more interested in the variability of loss rather than its mean value, which does not usually capture changes in network behaviour. Here, we use our previous works on the shift of interest in resilient networks design [5–7] as a basis. There, we discuss using business-related risk measures, such as Value-at-Risk (*VaR*). This measure is already being used in communication network failure descriptions [8], and security management [9]. Nevertheless, recurrent reservations exist about the fact that in the financial sector *VaR* is known to be misleading, due to its lack of the property known as *subadditivity*. As such, it is postulated to investigate more complex measures. The aim of this paper is first to show that these disadvantages do not appear if *VaR* is used to assess risk related to random network failures. We show

that only extremely atypical network characteristics, not met in practice, can disrupt the usefulness of *VaR*. Further, we provide a computationally effective method for estimating the level of penalties with this metric.

From the viewpoint of pay-offs, our results ensure: (a) effective quantification of total penalties imposed on a network operator due to failure presence, and thus (b) opening of a broad range of possible risk response methods based on *VaR* and elaborated in the financial field. The former enables an operator to save money in comparison with using a simple addition of risk measures calculated for a single service. Such an approach takes advantage of the diversification typical for investing [10]. The proposed estimation method is based on offline calculations, thus making them easier, more robust and less draining than simulations. The model calculates the penalties for both unprotected and protected connections. Our approach deals with a broad range of mappings between technical loss and business impact.

The remainder of the paper is organized as follows. First, we discuss related work in Sect. 2. In Sect. 3, we focus on various methods of expressing the monetary impact of failures (compensation policies) and methods for meaningful quantification of the predicted financial losses (risk measures). The section also discusses the known reservations concerning the main financial risk measure. In Sect. 4, we elaborate on an effective method for quantifying the upper bounds on this measure in the context of transport network services. This is the main contribution of our paper, and we show how to find the bounds for various compensation policies effectively. Then, we present numerical results confirming validity of our statements and models. In Sect. 5.1, we show that from a practical viewpoint the mentioned concerns on lack of subadditivity are not relevant in resilient network design, and thus it is possible to use *VaR* with all its advantages. In Sect. 5.2, we present the results of a very broad numerical study proving that the provided models for bounds of *VaR* are indeed exact. The final section concludes our work and shows avenues for future research.

## 2 Related Work

Franke [11] notes that the discussion of the relationship between the technical aspect and the business context related to network management is poorly developed. Our paper is presented in order to change this and boost work towards the goal of efficiently interfacing the technology and business worlds. While the methods and protocols for network resilience, described for instance in [12, 13], are not a new topic, many problems remain in a business-oriented approach to resilient network design and management. Historically, in the communications sector, risk has been dealt with for example in: (a) selection of new investment [14], (b) security against faults generated by malicious behaviour [15–18], or (c) quantification of deviations from the desired quality levels [19]. Risk assessment is the most popular concern researched in these contexts. While Value-at-Risk was postulated to be used in communications networks for risk quantification in network security [20, 21] and resilience [8, 9], there are no efficient theoretical models to predict *VaR* when it is applied in network resilience. At the modelling level, we use some elements of the

methodology similar to more recent works on network reliability, such as [22–25], yet we also deal with the modelling of whole distributions, and we add penalty [26], SLA [27, 28], and compensation policy [8] concerns.

Our numerical study is based on the distributions of failure and recovery times reported in literature. The relevant bibliography is reviewed and commented on in the context of risk engineering in [29]. Extensive studies of failure and recovery times related to operational networks have been performed and reported for the Sprint network [30], the Finnish research network [31], and the Norwegian university network [32]. Generally, a typical approach in numerical studies is to assume that the failures arise due to the homogeneous Poisson process. This means that times between failures are exponentially distributed. This classical approach seems to be statistically valid for many cases in communications [32, 33]. Other distributions for failure times, or their approximations by times between consecutive failures, are also occasionally reported (e.g., Weibull distribution [30]), although they cannot be responsible for generating heavy tails in the loss distributions. Modelling of downtimes (repair or recovery times) is more controversial. While the simplest approach also uses exponential times, recovery times in real networks appear to be log-normal [34] or Pareto-like—but always having mean value [31, 33].

This contribution can be treated as an extension of the two previously published papers [6, 7]. The new contribution can be summarized as follows: (1) The first paper [6] elaborates on the issue of lack of subadditivity of  $VaR$ , and shows that this risk measure can be successfully used in communications since lack of subadditivity is not a concern in practice. Here, we extend the set of the numerical studies to confirm this statement, especially by broadening the set of investigated distributions and taking into account node failures. Therefore, while the final statement is the same, now we increase the confidence on validity of this statement. (2) As the main contribution of this paper, we conceive the extension of [7], where the model for finding the upper bound on risk measures related to random failures in communication networks is presented. Our extension presented here is considerable: (a) we have extended the set of compensation policies to the ones that are more realistic than the ones shown in [7]; (b) while [7] uses a simple model based on results elaborated in a seminal work [35], here we present a more general model based on results derived in [36]; (c) the presented numerical studies are much broader than the ones presented before.

### 3 Business-Related Risk Assessment

SLA defines the desired values of parameters related to the services provided. These parameters include non-functional properties, such as reliability in the presence of network failures, the maximum acceptable downtime, or interval availability for a period of time [27, 37]. Penalties for not meeting these requirements may also be agreed and form the basis for calculating monetary impact to quantify business risk [26].

### 3.1 Compensation Policies

The way a penalty is defined as a function of the technical reliability parameter is known as the *compensation policy* [38]. Basically, if an outage appears and lasts for period of time  $\tau$ , we can model  $p$ —the penalty (outage cost) for this single outage. It is as a general function of  $\tau$ :  $p = f(\tau)$ . To find  $p$ , we follow the basic compensation options considered in [11, 38]. All of them are represented by convex functions. They are illustrated in Fig. 1.

It is possible to base the compensation policy on the number of all outages perceptible at the service level over a given interval. This means that the emphasis is put on the service *continuity*. This concerns services such as very short communication connection or sensitive data for real-time traffic control. Such services are rendered useless, no matter how brief outages are or how fast the resilience procedure is. We call such a policy *Cont*. For it, we may assume the fixed penalty independent of the outage time  $\tau$ :

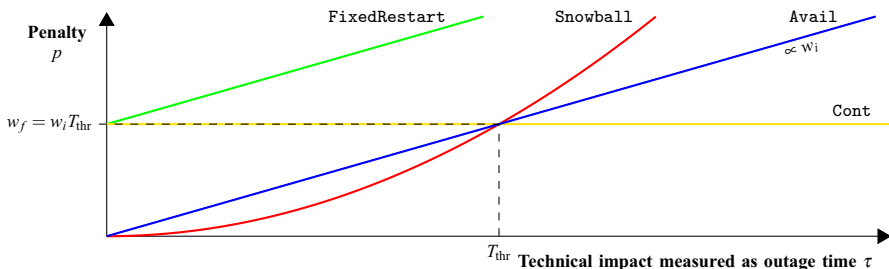
$$p_{\text{Cont}} = w_f. \quad (1)$$

Another compensation policy is to assume that the penalty is proportional to the downtime [39], i.e., the amount of time when the service is not operational. If the penalty is proportional to this time, this policy can be expressed by the so-called *interval unavailability*, the fraction of time when the service is not operating [40]. As this measure is a probabilistic complement to a better known availability, we call this policy *Avail*. This type of policy is most suitable for long-lasting services for elastic traffic, such as data transmission, web browsing, e-mail, and is typically agreed with individual customers. This reference policy based on the downtime  $\tau$  can then be expressed as:

$$p_{\text{Avail}} = w_i \tau. \quad (2)$$

The *Cont* and *Avail* policies can be combined arbitrarily. In the simplest case, we can use the fixed restart policy (*FixedRestart*), where additionally to the linear penalty due to the outage time, a fixed penalty  $T_{\text{thr}}$  for each outage is added:

$$p_{\text{FixedRestart}} = w_i(\tau + T_{\text{thr}}). \quad (3)$$



**Fig. 1** Compensation policies considered in the proposed model

To make things more complex, but close to reality, it can be assumed that the penalty is based on the number of outages exceeding a selected downtime *threshold* [41]. A practical example of this policy can be seen for SLAs related to Amazon S3, an online storage service.<sup>1</sup> Such a policy is valid when the agreement predicts the value of the Recovery Time Objective (RTO) beyond which the client can be sure that the failure will inflict some harm. This is a typical approach in business continuity planning [42].

Additionally, the penalty may not be scaled proportionally over a period of outage [43]. For instance, the so-called snowball effect [11] can be modelled by a compensation policy, where a strongly non-linear function explicitly shows the penalizing effect of a large value of the cumulative downtime or a number of outages. Such an approach may be especially relevant for relatively short services carrying sensitive data tolerating only short outages, such as transfers with advance reservation (e.g., in grid networks) or traffic with strict response limits. In our work, this situation is modelled with the non-linear smooth *Snowball* penalty:

$$p_{\text{Snowball}} = \frac{w_i}{T_{\text{thr}}} \tau^2, \quad (4)$$

where long downtimes are penalized considerably, while for the downtimes smaller than  $T_{\text{thr}}$  the penalty is sub-linear (or even negligible). For given values  $w_i$  and  $T_{\text{thr}}$ , the outage cost for  $\tau = T_{\text{thr}}$  is assumed to be the same in the cases of the *Avail* and *Snowball* policies. Hence, we obtain a form of ‘normalization’ of penalty values, as can be seen in Fig. 1.

### 3.2 Probabilistic Risk Measures

The values of penalties defined on the basis of a given compensation policy are random in a given time interval. To define meaningful quantification on their basis, so-called *probabilistic risk measures* are used. From the risk evaluation viewpoint, in the best case scenario the full probability distribution function (PDF) of the impact expressed in monetary units can be found. For a given PDF of the penalty, point estimates are applied [44]. The popular mean value of the penalty distribution [45, 46] has been found to be insufficient. The reason is its inability to quantify extreme values, characteristic of dealing with the risk context. Instead, the main measure used by finance departments to quantify the level of investment is Value-at-Risk, *VaR*. The definition is as follows [47, 48]:  $VaR_\eta$  is a quantile measure, and it provides for a selected level of probability  $\eta$  the value of penalty that can appear. Let  $\xi$  be the level of penalties to be paid in an interval. If  $P_\xi(x) = \Pr\{\xi \leq x\}$  is the cumulative distribution function of  $\xi$ , the Value-at-Risk is defined as the maximum penalty with a given confidence level  $\eta$ :

$$VaR_\eta = \inf\{x : \Pr\{\xi \leq x\} \geq \eta\} = P_\xi^{-1}(\eta). \quad (5)$$

<sup>1</sup> According to the most recent *Amazon S3 Service Level Agreement* (version: September 16, 2015; source: <https://aws.amazon.com/s3/sla/>), the outage starts to be counted if the requests are not responded for no less than 5 min.

This measure is now widely used in the banking sector to assess the obligatory level of savings. Additionally, it has been suggested that it is introduced in network design [8, 9, 29, 49, 50]. In response to some disadvantages of *VaR*, especially the lack of subadditivity (see Sect. 3.3 below), the derivative measures were proposed. However, although new methods for their efficient assessment have recently appeared [51, 52], in many cases their values are still at least one level of magnitude higher [6]. Thus, they cannot be treated as an efficient basis for investments and network design decisions.

### 3.3 Concerns about Subadditivity of *VaR*

There are some postulates to characterize *coherent* (desirable) risk measures [53]. One property is especially problematic when dealing with *VaR*. The property is known as *subadditivity*. It can be defined as follows:

$$\rho\left(\sum_i x_i\right) \leq \sum_i \rho(x_i), \quad (6)$$

where  $\rho$  represents the measure of risk for items  $x_i$ , e.g., services.

Subadditivity has the following positive consequences. (1) Quantification of risk measures is easier during the risk evaluation phase, where *risk aggregation* (calculation of the overall risk from individual risks) is conducted. In the case of subadditive measures, it is possible to easily assess the upper bound of risk. (2) Portfolio diversification, justifying good practice to provide service differentiation, is advantageous in comparison to separate investments [54]. (3) Avoiding or mitigating the risks related to the greatest levels of impact is the best option to deal with risk response [9]. (4) Subadditivity is a necessary condition to ensure the convexity of the risk measure. Then, efficient linear programming-based methods inspired by portfolio optimization approaches [44, 55] become feasible during network design.

A lack of subadditivity does not only mean that the above advantages are not present, but also that there is a very important danger related to using such a risk measure. It is believed that one of the roots of the banking crisis in 2008 was related to improper assessment of credit risks [56], i.e., based on *VaR*. The problem is that the used method is very sensitive to heavy tails in the PDFs of the impact. This sensitivity is the result of a lack of subadditivity. While it is common practice in the investment sector to base the *VaR*-related calculations on normal distributions [57], this is not always justified. These arguments against *VaR* are repeated in risk studies. Hence, we decided to verify whether lack of subadditivity is a real danger in resilient networks. This is done by numerical simulations in Sect. 5.1 and the results are very optimistic. Therefore, we assume to be able to use this risk measure without dangers.

The next part of the paper, then, focuses on the main contribution, which is an efficient modelling of *VaR* in resilient networks.

#### 4 Probabilistic Assessment of Aggregated Risk (Total Penalty)

Here, we assess the value of a risk measure for the total penalty paid by the network operator during a given time interval. One could think of a simple method of calculating the total penalty by calculating risk measure values for single services and then adding the results to obtain the aggregated risk measure related to the whole portfolio of services. This is not the best method, since it may provide an over-optimistic bound in the case of  $VaR$ , which is not subadditive. Additionally, even if subadditivity is observed as shown in Sect. 5.1, the bound obtained in this manner may be too pessimistic. This phenomenon is also noted later in our numerical results presented in Sect. 5.2. Therefore, we need an effective method of providing an upper bound for aggregated risk.

The following mathematical framework enables us to express the compensation policy if it is consistent among all the services. A network is represented by a graph  $(V, E)$ , where  $V$  is a set of nodes, and  $E$  is a set of links connecting the nodes.  $V \cup E$  is the set of network components. All of them are unreliable, which means they may fail and be repaired. Hence, we associate two probability distribution functions with each unreliable element: (a) the first describes time between failures and (b) the second concerns downtimes. While we use various types of time distributions (exponential, Weibull, Pareto, and log-normal), at some stage of the modelling we need to determine the failure ( $\lambda_c$ ) and repair ( $\mu_c$ ) rates for each unreliable network component  $c \in (V \cup E)$ . All the failure and repair processes are assumed to be independent of each other. Each service  $d$  is modelled at the physical level as a transfer service between two different nodes using a connection made of an  $n$ -tuple of links and nodes. The algorithm follows the steps given below:

1. Before we start preparing an exact model of risk, we need to assign compensation policies to determine penalties for each service. We also assume that each service is given a pre-determined primary path, and a backup path if the dedicated protection case is modelled. That is, we do not bother about the routing which is treated as an input to our algorithm.
2. Then, the continuous-time Markov chain (CTMC) for each service is constructed. Means and variances of compensation policy-related penalty values for all the services are found using these Markov chains. This makes it possible to find the *mean* and *variance* of the total aggregated penalty ( $p_{\text{Total}}$ ) over the interval.
3. Finally, the whole distribution of the aggregated penalty parameterized by these two values is found. We use one of the elliptical distributions. We found that, typically, the log-normal distribution gives the best fit results. When the whole risk distribution is parameterized, it becomes possible to find its quantiles, including  $VaR$ .

The different elements of this scheme are described below.



#### 4.1 General Case

To quantify a quantile risk measure (such as  $VaR$ ), we need to estimate a full probability distribution function for the penalty over a given time interval (typically *per annum*). We need to evaluate this value on the basis of the penalties calculated for separate services instead of modelling the level of penalties for the whole network, which would be a very complex task. The individual penalties related to various services are correlated because a failure of one component can affect many services. To estimate the level of the total penalty  $p_{\text{Total}}$ , we want to use the worst case approximation (the upper bound) for finding covariances between penalties calculated for various services.

First, let  $\mathbf{X} = [X_1, \dots, X_d, \dots]$  be the random vector of penalties for each service  $d$ . The total penalty, used to measure the aggregated risk, is calculated for an interval as:

$$p_{\text{Total}} = \sum_d X_d. \quad (7)$$

Then, let  $N_d(t)$  denote the number of outages that happen to service  $d$  during the observation interval  $t$ . And let  $p_d$  denote the single penalty (for a single outage) related to this service. The modelling of various types of penalties related to different compensation policies, and methods of finding penalty values, are discussed in Sect. 3.1. The penalty related to each service can be found as a random sum of individual penalties related to this service (i.e., for various outages). If we assume that the means and variances of  $N_d(t)$  and  $p_d$  are known, we can use basic probabilistic rules to find the average value of the total penalty over an interval for a service as:

$$E[X_d] = E[N_d(t)]E[p_d], \quad (8)$$

and its variance as:

$$D^2[X_d] = E[N_d(t)]D^2[p_d] + E^2[p_d]D^2[N_d(t)]. \quad (9)$$

Then, the average value of the total penalty can be found with the vector of average values of the penalties  $E[\mathbf{X}] = [E[X_1], \dots, E[X_d], \dots]$  as:

$$E[p_{\text{Total}}] = E[\mathbf{X}]\mathbf{1}^T, \quad (10)$$

where  $\mathbf{1}$  is the vector of ones  $[1, 1, \dots, 1]$  of the appropriate dimension (here, of length equal to the number of services). The variance of the whole penalty is found with the help of the covariance matrix of  $\mathbf{X}$ :

$$D^2[p_{\text{Total}}] = \mathbf{1}^T \text{Cov}(\mathbf{X})\mathbf{1}. \quad (11)$$

Finally, these values are used to find the parameters of an elliptical distribution being the distribution of a total sum of penalties, or a distribution close to it—for an explanation, see [35]. In our case, this is the log-normal distribution. On the basis of

this distribution, we are finally able to find the quantile risk measures. However, the problem is finding the values for parameterization. As exact calculation is too complex in practice, here we present the upper bound of the distribution to effectively obtain the values.

## 4.2 Unprotected Case

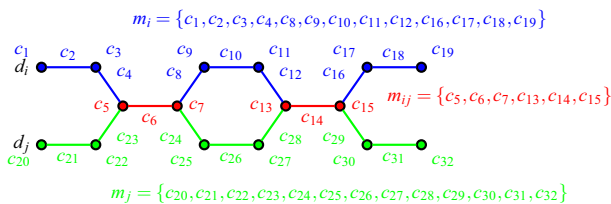
To show the idea of upper bounding, let us begin with a simple example of a service supported by an unprotected connection. Such a connection can be treated as a series reliability structure that fails when there is a fault in at least one of its elements. To consider the correlation between the events, let us consider two services  $d_i$  and  $d_j$ , each with its own SLA inducing specific penalty values. In our model each service uses the same type of compensation policy, although the related weights differ. For two services  $d_i, d_j$ , we define the following partition for three sets of all the components of the carrying connections (see Fig. 2):  $m_i = \{c : c \in d_i \wedge c \notin d_j\}$ ,  $m_j = \{c : c \notin d_i \wedge c \in d_j\}$ , and  $m_{ij} = \{c : c \in d_i \wedge c \in d_j\}$ . We call these sets ‘macrocomponents’. If we find the respective penalties  $X_{d_i}$  and  $X_{d_j}$  for the services  $d_i$  and  $d_j$ , the following bounds hold:

$$\begin{aligned} X_{d_i} &\leq X_{m_{ij}} + X_{m_i}, \\ X_{d_j} &\leq X_{m_{ij}} + X_{m_j}, \end{aligned} \quad (12)$$

where  $X_{m_{ij}}$  is the penalty calculated for the service consisting of macrocomponent  $m_{ij}$ , the case when  $d_i$  and  $d_j$  were simultaneously down.  $X_{m_i}$  and  $X_{m_j}$  are the penalties found for the virtual services consisting of macrocomponents  $m_i$  and  $m_j$ , respectively. The given inequalities come from the fact that there can be simultaneous failures in  $m_i$  and  $m_{ij}$ . Additionally, we were able to formulate these inequalities for penalties, while in fact they hold for outage times. However, on the basis of Jensen’s inequality we know that Eq. (12) are true, due to the fact that we use convex penalty functions. Then, using basic probabilistic rules, it is easy to show that for the upper bounds we have:

$$\text{Cov}[X_{d_i}, X_{d_j}] = D^2[X_{m_{ij}}] \quad (13)$$

if only  $X_{m_{ij}}$ ,  $X_{m_i}$ , and  $X_{m_j}$  are independent.



**Fig. 2** Illustration of partitioning to macrocomponents for connections supporting two unprotected services  $d_i, d_j$

We can treat a macrocomponent as just an ensemble of independent ON-OFF components. As such, it can be modelled as a single ON-OFF system itself. However, the general modelling of such an ensemble on the basis of the behaviour of a single element is not possible unless the uptimes and downtimes are assumed to be exponentially distributed. Hence, each component and the entire ensemble can be modelled as a CTMC. Then, it is possible to find the analytical bound for the risk measures. In practice, network failures arrive according to a Poisson process [31]. This is a common assumption taken while the mathematical modelling of failures is performed. In our numerical studies shown in Sect. 5.2, we challenge this assumption and show that our model also provides useful results when the failure process is not memoryless.

Let the state of a single component be modelled as a CTMC on the state space  $\{0, 1\}$  with failure rate  $\lambda_i$  and repair rate  $\mu_i$ . The state of the macrocomponent being an ensemble of  $n$  independent components is a CTMC on the state space  $\{0, 1\}^n$  with an infinitesimal generator matrix:

$$\mathbf{Q} = \bigoplus_{i=1}^n \mathbf{Q}_i \quad (14)$$

where  $\mathbf{Q}_i$  is the generator of a single component [58]:

$$\mathbf{Q}_i = \begin{bmatrix} -\lambda_i & \lambda_i \\ \mu_i & -\mu_i \end{bmatrix}, \quad (15)$$

and  $\bigoplus$  denotes the Kronecker sum. The above follows directly from the properties of the Kronecker product [59] and the independence of individual components. Then, we are able to find the bound related to Eq. (13).

As a macrocomponent is a series reliability structure, we are able to define only ensemble modelling CTMC, where all the components operate. Therefore, the time the system spends in the up-state ( $U$ ) is exponentially distributed with the rate being equal to the sum of failure rates of all the components. On the other hand, the distribution of downtimes is more difficult to compute. We use the embedded Markov chain and the Laplace transform to find a good approximation for the mean and variance of the time the system spends in the down-state ( $D$ ). Let us begin with a fully operational macrocomponent (all up-states). Next, a component fails after an exponentially distributed time. In the next Markov chain jump, the failure may be repaired or another failure may occur. The time elapsed before the next event is again exponentially distributed, with the rate parameter dependent on the current state. Applying the total probability formula to the number of failures, the distribution of  $D$  can be expressed as an infinite sum of convolutions. In the Laplace transform domain, convolutions become multiplications, and the first and second raw moment of the distribution can be derived from the derivative of the transform [58]. Finally, since the probability of simultaneous multiple failures is extremely low, the infinite series can be approximated by the first few terms, where in practice the first two to three terms are sufficient. This truncation simply omits possibility of triple, quadruple, or more simultaneous failures.

Below, we generalize the case of an unprotected connection outlined above to the case when the service can be supported by a more complex connection, especially

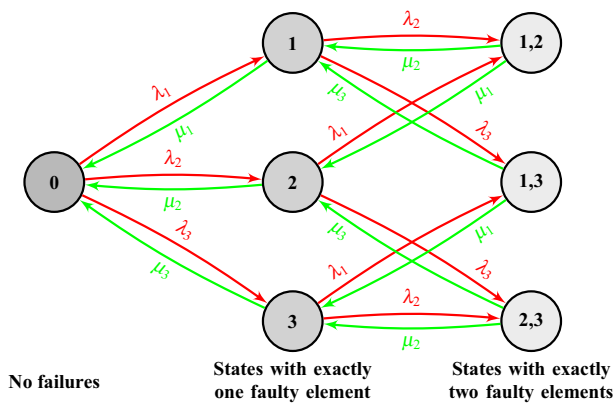
applying alternative connections for dedicated protection. Then, we expand the concept of a macrocomponent and relate it to the probability process state. The generalization is necessary since the model given in Eq. (14) is prone to the state space explosion. Despite the current computational power offered by an efficient sparse matrix implementation (as available in MATLAB, for instance), it is necessary to reduce the complexity of the model by redefining the state space.

### 4.3 Generalized Markov-Based Modelling of Penalties

The new and generalized state space is defined as follows:

$$S = \bigcup_{k=0}^M C_n^k, \quad (16)$$

where  $C_n^k$  is a set of all  $k$ -element combinations out of all  $n$  unreliable network components related to a given service (i.e., we deal with the components forming the working path for unprotected connection or a pair of paths for the protected case). And  $M$  is the arbitrary selected number of probable simultaneous failures in the network. In practice, we can assume that  $M \leq 3$  is sufficient, and the dimension of the state space is reduced from  $2^n$  to a number  $\sim n^M$ . Note that for  $M = n$ , both spaces (i.e., the one defined here as a generalization and the one related to Eq. (14) shown for the unprotected case in Sect. 4.2) are isomorphic. The new representation allows us to cut off configurations with extremely low probabilities by reordering original states. Here, we rely on the simple fact that having three, four or more simultaneous failures in a network is extremely unlikely. The infinitesimal generator of the process is a sparse matrix containing the entries of the following form (see example for  $n = 3$  and  $M = 2$  in Fig. 3):



**Fig. 3** An example of the CTMC when there are three ( $n = 3$ ) components in the network; we assume that there are no more than two faulty ( $M = 2$ ) elements at a time.  $\lambda_i$ : failure rate of element  $i$ ;  $\mu_j$ : repair rate of element  $j$

$$\forall_{i \neq j} \quad Q_{ij} = \begin{cases} \lambda_k & \text{if } s_j = s_i \cup \{k\} \\ \mu_k & \text{if } s_i = s_j \cup \{k\}, \quad s_i, s_j \in S. \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

The diagonal elements are calculated on the basis of the sums of rows:

$$Q_{ii} = - \sum_{j=1}^n Q_{ij}. \quad (18)$$

Note that since the state space is reduced in comparison to the full space related to Eq. (14), the exact formulas can be used instead of approximations proposed in [7] on the basis of the results found by Takács [35].

Now, for each service  $d_i$  we define only a pair of macrocomponents  $(U_i, D_i)$  constituting the partitioning of the whole set of states:  $S = U_i \cup S_i$ ,  $U_i \cap S_i = \emptyset$ . In some states  $U_i \subseteq S$ , the service  $d_i$  works, while in others summarized as  $D_i = S \setminus U_i$  this service is down. Note that the selection of  $(U_i, D_i)$  is unique for each service  $d_i$ . For instance, this is the case of an unprotected connection service,  $U_i = \{S_0\}$ , where  $S_0$  is the only state where all the components  $c \in d_i$  work properly. However, we keep this derivation general, so it remains useful for the dedicated protection case. In the latter case,  $U_i$  is the set of all states in which all the components of the primary path or all the components of the backup path are operational.

The calculation of the downtime of a macrocomponent simply involves solving the first *passage time* problem in CTMC, that is the amount of time it takes for the Markov process to reach the absorbing state from the initial totally faultless state [36] (that is, the time it takes for a system to jump out of  $U_i$  to  $D_i$ ). Although we use the CTMC, meaning the times between changing the various states are exponential, the overall passage time it takes to reach the macrocomponent  $D_i$  from  $U_i$  is not exponential. Therefore, if we wish to describe the state of the service on state space  $\{0, 1\}$  (0: the whole service is operational, 1: the service is faulty), we need to use a semi-Markov process with phase-type distributed sojourn times, which are the times the underlying CTMC spends in a groups of states  $U_i$  and  $D_i$ .

In order to find the distribution of sojourn times in  $U_i \in S$ , the states  $D_i = S \setminus U_i$  are assumed to be absorbing and the general formula for the passage time is found on the basis of the moments of time distribution, using the Laplace transform. It is found by rearranging the generator  $\mathbf{Q}$  defined by Eqs. (17)–(18). We know that it is possible to partition  $\mathbf{Q}$  to the following submatrices [36]:

$$\mathbf{Q} = \left[ \begin{array}{c|c} -\mathbf{q} & \mathbf{q}^T \\ \hline \mathbf{r} & \mathbf{T} \end{array} \right]. \quad (19)$$

The partitioning assumes the following state reordering:  $U_i$  is formed of the first  $r$  states. The values in the square matrix  $\mathbf{r}_{|U_i| \times |U_i|}$  are related to the transitions between all the states in  $U_i$ . Bearing this in mind, the matrices  $\mathbf{T}_{|D_i| \times |D_i|}$  (gathering data about transition rates inside the  $D_i$  macrocomponent) and  $\mathbf{q}_{|D_i| \times |U_i|}$  (gathering data about

transition rates between the states from various macrocomponents) are found in a unique way.

Now, the Laplace transform of the first passage time can be expressed as follows [36]:

$$\mathcal{L}(z) = \mathbf{P}_{\text{in}}(z\mathbf{I} - \mathbf{T})^{-1}\mathbf{r}\mathbf{1}, \quad (20)$$

where  $z$  is the complex variable defined for the Laplace transform, and  $\mathbf{I}$  is the identity matrix of the proper dimension. Additionally,  $\mathbf{P}_{\text{in}}$  is the distribution of the initial state of service  $d_i$ : if  $U_i \neq \{S_0\}$ , there are different starting points in  $U_i$  as well as in  $D_i$ , distributed according to  $\mathbf{P}_{\text{in}}$ . Since we are interested in distributions of the states upon the state change, we approximate  $\mathbf{P}_{\text{in}}$  by the stationary distribution of the embedded Markov chain of  $\mathbf{Q}$  conditioned on being in the selected state subset  $U_i$ . This way, we avoid solving differential equations to find the exact form of  $\mathbf{P}_{\text{in}}$ .

Now we are able to find the aggregated penalty values when various compensation policies are used. First, consider the raw moment of the distribution of the total penalty  $p = f(\tau)$ . By using the approximation of  $f$  with the Taylor series, the moments can be expressed in terms of the raw moments of the distribution of outage time  $\tau$ . The  $i$ th moment of the  $\tau$  distribution, denoted as  $m_n$ , can be found as follows [36]:

$$m_n = (-1)^{n+1} n! \mathbf{P}_{\text{in}} \mathbf{T}^{-(n+1)} \mathbf{r}\mathbf{1} \quad (21)$$

In this way, we can obtain values for non-linear compensation policies. For example, for the `Snowball` policy, we have:

$$\mathbb{E}[p_{\text{Snowball}}] = w_i \frac{-2\mathbf{P}_{\text{in}} \mathbf{T}^{-3} \mathbf{r}\mathbf{1}}{T_{\text{thr}}}. \quad (22)$$

The equation is obtained from the mean value of the penalty defined in Eq. (4), where the mean is the first moment  $m_1$ . It is then possible to replace the outage time  $\tau$  value by its moment calculated on the basis of Eq. (21). Similarly, using the relationships between various raw moments, it is possible to find the variance for this compensation policy:

$$\mathbb{D}^2[p_{\text{Snowball}}] = w_i^2 \frac{-24\mathbf{P}_{\text{in}} \mathbf{T}^{-5} \mathbf{r}\mathbf{1} - \mathbb{E}[p_{\text{Snowball}}]^2}{T_{\text{thr}}^2}. \quad (23)$$

Here, we take advantage of the fact that the fourth moment of a random variable is the second moment of its square.

#### 4.4 Dedicated Protection Case

As stated, the concept of macrocomponents is useful for the unprotected case. However, in the case of dedicated protection, macrocomponents—as understood according to the definition in Sect. 4.2—are coupled and cannot be modelled independently. State subsets  $U_i$  and  $D_i$ , defined as macrocomponents in Sect. 4.3, cannot be constructed from the common macrocomponent only. Then, the system of

the working and backup paths has to be modelled as one, and appropriate states are selected for  $D_i$ . Moments of the distribution of the number of faults of a service over an observation interval  $t_{\max}$  ( $N_d(t_{\max})$ ) must be calculated from the moments of a distribution of the passage time (i.e., the time to hit the set  $D_i$ ), since the stream of failure events observed by the service  $d$  does not form a Poisson process. Then, they have the form [35]:

$$E[N] = \frac{t_{\max}}{m_{\text{on}_1}}, \quad (24)$$

and:

$$D^2[N] = \frac{t_{\max}(m_{\text{on}_2} - m_{\text{on}_1}^2)}{m_{\text{on}_1}^3}, \quad (25)$$

where  $m_{\text{on}_i}$  is the  $i$ th moment of the up-time of the connection supporting service  $d_i$ .

When it comes to covariance in the case of dedicated protection, it becomes even more complicated, since the system is constructed from four paths (i.e., two primary and two backup paths). In such a system, the new subsets for services  $d_i$  and  $d_j$  are selected as follows:  $D_{d_{ij}} = D_i \cap D_j$  and  $U_{d_{ij}} = S \setminus D_{d_{ij}}$ . Then, it holds that:

$$\text{Cov}[X_{d_i}, X_{d_j}] = D^2[X_{d_{ij}}], \quad (26)$$

where  $X_{d_{ij}}$  is the penalty computed for the set of states  $D_{d_{ij}}$ . The reasoning is the same as that presented in relation to Eq. (13).

## 5 Numerical Studies

First, we would like to show that in practical cases, *VaR* behaves as if it was subadditive. We will show that the lack of subadditivity is seen only for the stochastic parameters that are not met in reality in networks. Next, we show that the theoretical model for risk assessment gives a very good bound.<sup>2</sup> The numerical examples are constructed as follows. The network topologies used are retrieved from the SNDlib library (<http://sndlib.zib.de>) [60] and model two large networks: the compact and dense German Research Network (`nobel-germany.xml`), and the very broad yet sparse US Network (`nobel-us.xml`). For each node and link in a network, the interchanging failure and resilience process is modelled. In the basic case, according to the most commonly assumed conditions, both distributions for link/node failure times/downtimes are exponential. Their rates were taken from [31]. We use the following function to find the failure/repair rates for links:

<sup>2</sup> Due to the limited space, we are not able to present in this paper all the obtained results. Therefore, they are present in the form of plots in the companion webpage: <http://home.agh.edu.pl/~cholda/research/effective-risk-assessment-with-value-at-risk/>.

$$\left\{ \begin{array}{ll} \lambda = \lambda_R \wedge \mu = \frac{1.3}{2.3m} & \text{if } l < 25 \text{ km} \\ \lambda = \lambda_R + 9 \frac{\lambda_R l}{l_{\max}} \wedge \mu = \frac{1.3}{2.3 \left( m + 9 \frac{ml}{l_{\max}} \right)} & \text{if } l \geq 25 \text{ km} \end{array} \right. \quad (27)$$

where  $l$  represents a link length expressed in kilometres ( $l_{\max}$  is the largest link length in a networks), and  $\lambda_R$  and  $m$  are the basic distribution parameters retrieved from [31]. Each service has its own parameters necessary to find the exact value of the penalty. The scaling weight  $w_i$  is equal to the volume transferred by a service. This volume is provided with the network models. The time scale  $T_{\text{thr}}$  is equal to the mean downtime of the most reliable component of the network. We have checked that scaling this value with 0.5 or 2 does not change the qualitative character of the results. Each connection supporting a service is routed with the shortest path routing found by the Dijkstra algorithm (our networks are modelled by weighted digraphs, where the weights representing lengths of the links are non-negative). For each scenario, we held 100,000 simulations developed in C++. Each simulation time was 1 year of network operation; this is the interval for which penalties due to the assumed compensation policies are estimated. We need this number of simulations since the events are rare, and only with 100,000 simulations do the observed correlations for two runs start to differ at the third decimal place. The mathematical modelling is performed with the help of MATLAB.

A typical assumption made during various risk assessment calculations is that all the failures and repairs are independent and the downtimes are exponentially distributed. We support the former assumption. On the other hand, this is a rough estimate of the reality of the situation. Additionally, with exponential downtimes, we cannot observe the non-subadditive character of  $VaR$ , since heavy tails are not present. Additionally, it has been reported that PDFs of recovery times in networks can be heavy-tailed [30, 31, 33]. We use such distributions (the Pareto distribution), and by changing parameters, we show that the lack of subadditivity of  $VaR$  does appear for extremely atypical values only.

The following options for the simulation scenarios are defined:

1. Networks:
  - (a) German ( $N_{\text{Ger}}$ ),
  - (b) US ( $N_{\text{US}}$ ).
2. Resilience methods:
  - (a) unprotected service ( $R_{\text{UP}}$ ),
  - (b) dedicated path protection ( $R_{\text{DP}}$ ).
3. Compensation policies:
  - (a) Cont,
  - (b) Avail,



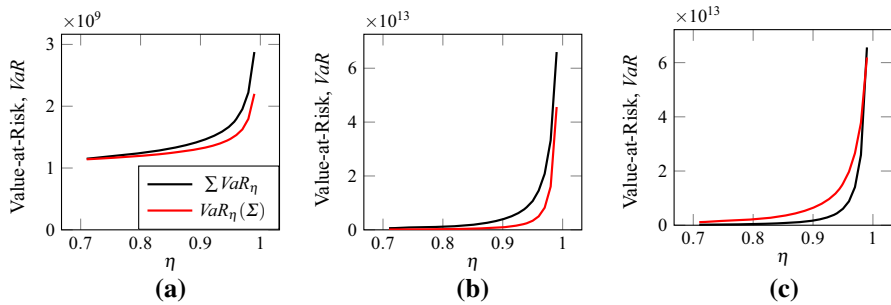
- (c) FixedRestart,
  - (d) Snowball.
4. Distributions of failure times and downtimes:
- (a) E: negative exponential  $\text{Exp}_\lambda (f_{\text{exp}}(t) = e^{-\lambda t})$ ;
  - (b) P: Pareto  $\text{Par}_{\alpha,m} (f_{\text{Pareto}}(t) = \frac{\alpha m^\alpha}{t^{\alpha+1}})$ , note that for this distribution there is no mean when  $\alpha \leq 1$  and no variance when  $\alpha \in [0, 2]$ , the  $m$  parameter is 60 sec in all cases (this value stems from the granularity of router queries sent by the Simple Network Management Protocol to check the connectivity state in the broad numerical study shown in [31]) except for the extreme case that attains non-subadditive  $\text{VaR}$ ;
  - (c) W: Weibull;
  - (d) L: log-normal.
5. Risk measures:
- (a) the value of the aggregated risk obtained in the simulation:  $\text{VaR}(\Sigma)$ ,
  - (b) naïve upper bound for aggregated risk:  $\sum \text{VaR}$ ,
  - (c) efficient upper bound introduced by the theoretical model presented in this paper:  $\text{VaR}_{\text{Th}}$ .

In all the cases,  $\eta$  means the quantile level.

### 5.1 Study I: Subadditivity of Value-at-Risk is Not an Issue

The general character of the obtained results is shown in Fig. 4, grouping the three representative cases. It is related to the US network, but the character of the results is the same for the German network. In this figure, only the results for the Avail compensation policy are presented, as it is most sensitive to the recovery time distribution. In the figure we show two curves. One is related to the risk measure calculated separately for each connection individually and summed (the naïve upper bound exceeded if the measure shows lack of sub-additivity). The other curve shows the value of the measure calculated for the distribution of the total penalty in the network. Figure 4a shows the situation when both failure and recovery times are exponential, and where the character of the PDF for the penalties is convergent to the Gaussian-, Gamma, or log-normal-like distribution, which is the result of the fact that the cumulative downtime distribution is the convolution of exponential times. Therefore, the subadditivity holds, and further on we do not analyze results for exponential recovery times.

The option to show subadditivity appears only when using heavy-tailed distributions. The second plot (Fig. 4b) shows the evident subadditivity of  $\text{VaR}$ , even though it is related to atypical and unrealistic parameters of the Pareto



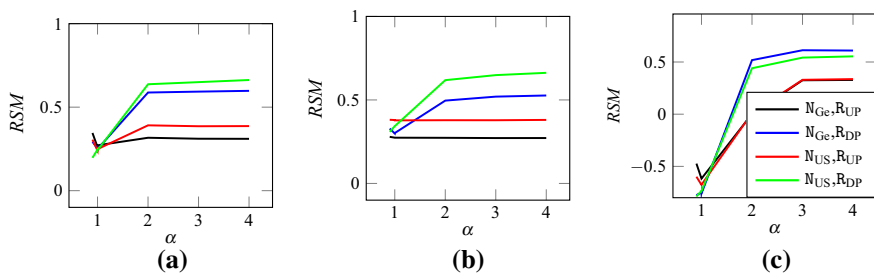
**Fig. 4** Examples of the risk parameters quantified with  $VaR$ . Network:  $N_{US}$ . Resilience method:  $R_{UP}$ . Compensation policy: Avail. **a** Distr. of fail. times:  $Exp_{1/90}$ . **b** Distr. of fail. times:  $Par_{3,60}$ . **c** Distr. of fail. times:  $Par_{0.9,1}$

distribution, where the penalties start to present heavy-tailed behaviour. The lack of subadditivity is not visible until we use extremely strange parameters, as shown in Fig. 4c. Even then, the subadditivity is present for high quantile values (e.g., 0.99) that are commonly used by the financial sector. The measure is not subadditive for lower quantile values, such as for 0.95 which may also be interesting in the application context. As it cannot be seen in the figures, where we show only large quantile values, it is also worth mentioning that the results also conform with the theory [53] that  $VaR_\eta$  based on the normal distribution of losses is subadditive for values of  $1 - \eta$  smaller than  $\frac{1}{2}$ . And indeed, we are interested in large values of  $\eta$ , which is why we study mainly quantiles for  $\eta > 0.9$ .

To show how the situation changes for all the studied Pareto-based simulation scenarios, we extended their number by changing the value of  $\alpha$  responsible for the existence or lack of a heavy tail. To present the results in a compact form, we introduce the metric known as *Relative Subadditivity Measure*, defined as:

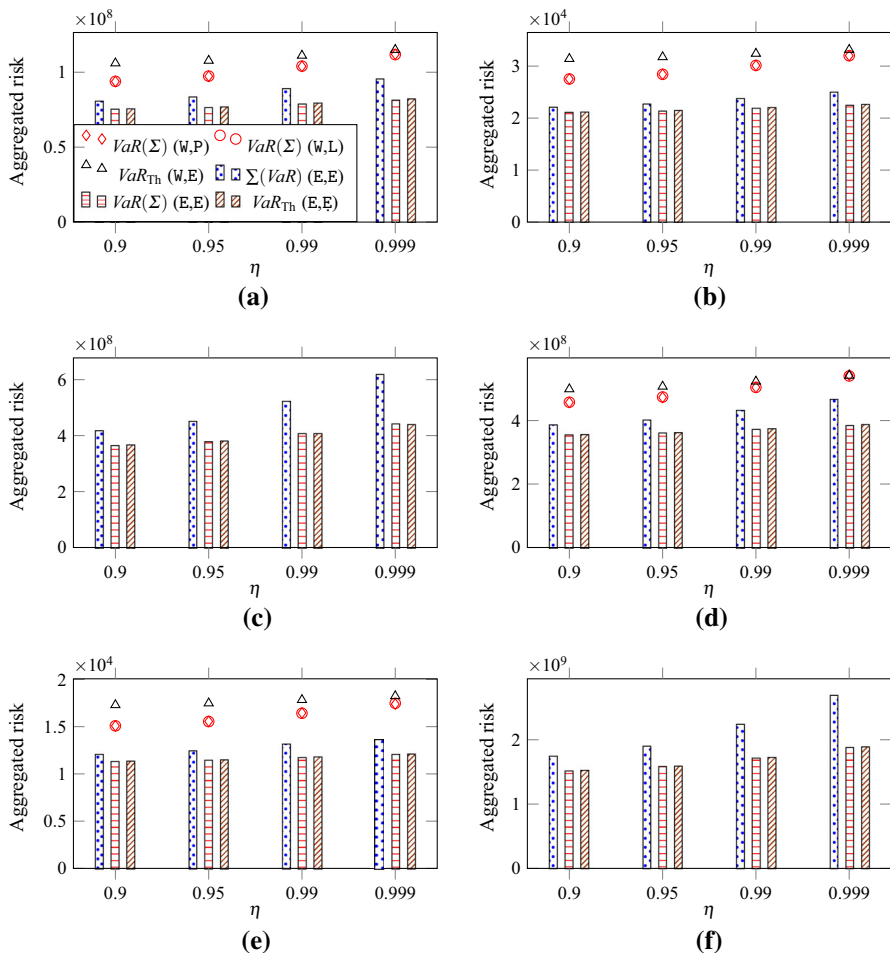
$$RSM = \frac{\sum VaR - VaR(\Sigma)}{VaR(\Sigma)}, \quad (28)$$

The non-negative values of  $RSM$  show the appearance of subadditivity. The results, grouped in three sub-figures related to compensation policies in Fig. 5, show that



**Fig. 5** Values of  $RSM$  calculated for  $VaR_{0.95}$  and simulation scenarios with distribution of failure times according to:  $Par_{2,60}$ . **a** Comp. policy: FixedRestart. **b** Comp. policy: Cont. **c** Comp. policy: Snowball

only the compensation policy emulating the snowball effect may show the lack of subadditivity. As already stated, this is only the case for highly unrealistic distributions. The results for the other compensation policies are loosely independent of the distributions. The level of subadditivity is somewhat dependent on the selected resilience method, while here it seems that the approximation of the aggregated risk measure by the sum of the individual risk values is better for unprotected connections. For protected connections, the upper bound becomes more pessimistic. Generally, we can see that *VaR* can be reliably used for assessing risk in all types of networks, with various compensation policies, and using different resilience



**Fig. 6** Results comparing the exact values of aggregated risk measures and the derived upper bounds for them when the connections are not protected. The legend presents the used distributions as  $(F, R)$ , where  $F$  represents the distribution of the time to failure, and  $R$  represents the distribution of the downtime. **a** Comp. policy: FixedRestart. Network:  $N_{Ge}$ . **b** Comp. policy: Cont. Network:  $N_{Ge}$ . **c** Comp. policy: Snowball. Network:  $N_{Ge}$ . **d** Comp. policy: Avail. Network:  $N_{US}$ . **e** Comp. policy: Cont. Network:  $N_{US}$ . **f** Comp. policy: Snowball. Network:  $N_{US}$

methods, unless the behaviour of the recovery times becomes highly atypical. The only case we can imagine, apart from wartime, is related to natural catastrophes. However, in such an event it is doubtful whether typical business-related risk management would be relevant at all.

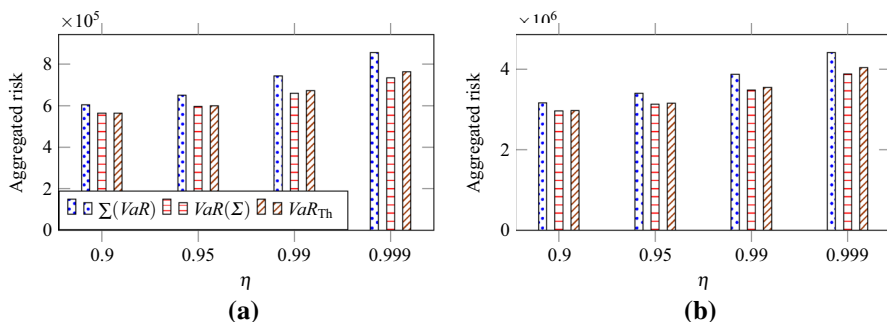
## 5.2 Study II: Effective Upper Bounding of Risk Measures

We know that, while calculating  $VaR$ , we should not use the summation of the values of  $VaR$  for individual services, as this measure is not subadditive in general. And even if it happens to behave as if it was subadditive, we would like to obtain more exact estimates. Therefore, we show that the theoretical modelling derived in Sect. 4 provides a very good upper bound for the aggregated risk, much better than the naïve bound obtained by simply adding the risk measure values across all the services separately.

As the Weibull distribution is treated as the most suitable to model time to failure [30], we decided to provide the heuristic selection of the exponential distribution to fit it. The Weibull uptime distribution is then approximated by the exponential distribution in the following way. We consider two simple ON-OFF systems. One with Weibull uptime distribution and exponential downtime distribution (we call this system ‘the Weibull system’), and the second with both exponential distributions (‘the exponential system’). The Weibull distribution is approximated by exponential distribution such that for both systems:

$$m_{\text{Weib}} + a\sigma_{\text{Weib}} = m_{\text{Exp}} + a\sigma_{\text{Exp}}, \quad (29)$$

where  $m_{\text{Weib}}$ ,  $m_{\text{Exp}}$ ,  $\sigma_{\text{Weib}}$ , and  $\sigma_{\text{Exp}}$  are the expected values and standard deviations of the total downtime during  $t_{\text{max}}$  for the Weibull and exponential systems, respectively. It was observed that the value  $a = 1.2$  gives quite accurate upper bound for values related to the total penalties. We use the same simulation settings as in the case of Study I presented in Sect. 5.1. After calculating the parameterized



**Fig. 7** Results comparing the exact values of aggregated risk measures and the derived upper bounds for them when the connections use dedicated protection. Compensation policy: FixedRestart. It is assumed that both the time to failure and downtime distributions are exponential. **a** Network:  $N_{\text{Ge}}$ . **b** Network:  $N_{\text{US}}$

log-normal distribution of penalties (this distribution is best fitted to evaluate the tail of the total penalty for most of the studied policies) for all the services, we are able to calculate  $VaR$  for each of them for the following quantiles:  $\eta = 0.9, 0.95, 0.99, 0.999$ . However, as we are interested in the estimation of the risk measure related to the total penalty ( $VaR(\Sigma)$ ) that should be paid by the operator, in Figs. 6 and 7 we do not present the values of the individual risk measures. We only focus on showing that the proposed upper bound ( $VaR_{Th}$ ) is much more effective than the upper bound obtained with the summation of risk measures calculated for separate services ( $\sum(VaR)$ ). We present only the most interesting and non-intuitively good results. For instance, the most prone to bad bounding are the systems with the Pareto downtimes, therefore—where possible—we show the performance for them. Due to complexity of calculations, the results shown for protections concern only ten services with the largest values of penalties agreed in their SLAs.

As can be seen from the figures, even if  $VaR$  in this case does not show a lack of subadditivity (i.e. advantageous effect of diversification appears), the bounds provided by summing the individual values of the measures are much more pessimistic than the worst case bound we elaborated. By summing the risk measures for the individual services, an overestimation of approximately 10 % is provided, while our upper bound gives an almost perfect fit with an average overestimation of only 0.10 %. Using the provided upper bound, we can estimate the exact value of the total penalty without needing to perform a very large number of simulations for all compensation policies. Such a large number is indispensable in order to obtain sufficiently small confidence intervals if the total penalty is to be obtained by simulation. The confidence intervals for quantiles diminish quite slowly, i.e., with the speed proportional to a value between  $(n^{-2}, n^{-1})$ , where  $n$  is the number of samples [61]. Moreover, for large values of  $\eta$ , if the tails of the distributions do not decay quickly, we have to increase the number of simulations to have the opportunity to calculate the quantiles to gather enough samples. Our approach enables us to avoid technical problems such as this. It is noticeable that for the Snowball compensation policy combined with the Pareto downtimes we are not able to effectively bound the results.

### 5.3 Summary of the Results

We promote the usage of business-relevant risk measures in the context of resilient networks. That is, we propose to apply the commonly accepted quantile measure  $VaR$ , which is widely used in the investment sector to assess the obligatory level of savings. In network design and management, this approach can be used to predict penalties, estimate the level of the provided protection against failures, or suggest necessary changes to the network. Nowadays, simpler measures of risk used in network design lose information about impact variability, since they are based on mean values. Therefore, they describe the character of the impact distribution very roughly. On the other hand,  $VaR$  preserves the information on variability and makes it possible to use complex portfolio optimization methods elaborated in the financial

sector. Nevertheless, we paid attention not only to the advantages in using this measure, but also highlighted potential drawbacks. We showed that: (1) For a broad spectrum of distributions encountered in real networks, the *VaR* measure is subadditive in practice and can be used reliably. (2) Additionally, even when we were able to find highly unrealistic values for which the lack of subadditivity is expressed, it was generally the case for low quantile values that are less interesting in practical cases of risk assessment. (3) Furthermore, as the quantification of *VaR* requires calculation of the whole penalty distributions, we propose a computationally effective method of exact upper bounding of the total penalty to be paid by the operator using various compensation policies encountered in practice. (4) While our newly introduced model assumes memoryless property of the involved stochastic processes, we show that it also performs well when challenged with various non-exponential distributions.

## 6 Conclusions

With the results confirmed experimentally and summarized in the previous section, we are providing the operators with a tool to assess the business consequences of technical losses, which will improve SLA preparation as well as network design and management processes. This may also be used for resilience purposes, e.g. selection of network parts to be especially protected. The calculated values can also be used in optimization problems, thus opening the possibility of using the methods elaborated in modern portfolio theory.

We regard making the most of this potential as future work, where we would like to: (1) Focus on mathematical programming-based optimization approaches that treat connections or service classes in resilient networks as investments, and where risk assesses either the return from selling them or the loss that is incurred when the services are lost due to failures. (2) Extend the presented model by relaxing the constraint on the unified type of compensation policy across all services. (3) Add relevant modelling for the shared protections, where the backup resources are not dedicated to selected services anymore. (4) Deal with the service which is not based on an end-to-end (unicast) connection, but is related to a connection to a pool of resources, where availability of only one item is sufficient to provide the service. From data transfer viewpoint, this scheme can be perceived as related to anycast. A practical application of this case is relevant to cloud or grid environments. While this is a problem of utmost practical importance, the related modelling is more complex than the one used by us in this paper. The reason for increased complexity is that the level of the service dynamics involved is much higher and the IT infrastructure is dependent on external resources, such as power provisioning. The latter involves hard problems known under the name of system-of-systems modelling.

**Acknowledgments** This scientific work was financed by the Polish Ministry of Science and Higher Education from the research budget for 2013–2015, Project No. IP2012 022972. This research was supported in part by PL-Grid Infrastructure.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Trček, D.: Computationally supported quantitative risk management for information systems. In: Gülpınar, N., Harrison, P., Rüstem, B. (eds.) *Performance Models and Risk Management in Communications Systems*, Springer Optimization and Its Applications, pp. 55–77. Springer, New York (2011)
- Banerjee, S., Shirazipourazad, S., Ghosh, P., Sen, A.: Beyond connectivity—new metrics to evaluate robustness of networks. In: *Proceedings of 12th IEEE International Conference on High Performance Switching and Routing HPSR 2011*, Cartagena, Spain (2011)
- Araujo Wickboldt, J., Bianchin, L.A., Castagna Lunardi, R., Granville, L.Z., Gaspary, L.P., Bartolini, C.: A framework for risk assessment based on analysis of historical information of workflow execution in IT systems. *Comput. Netw.* **55**(13), 2954–2975 (2011)
- Todinov, M.: *Risk-Based Reliability Analysis and Generic Principles for Risk Reduction*. Elsevier, Amsterdam (2006)
- Cholda, P., Jaglarz, P.: Optimization/simulation-based risk mitigation in resilient green communication networks. *J. Netw. Comput. Appl.* **59**, 134–157 (2016)
- Cholda, P., Guzik, P., Rusek, K.: Risk-awareness in resilient networks design: Value-at-Risk is enough. In: *Proceedings of 16th International Telecommunications Network Strategy and Planning Symposium NETWORKS 2014*, Funchal, Madeira, Portugal (2014)
- Cholda, P., Rusek, K., Guzik, P.: Upper bound for failure risk in networks. *Electron. Not. Discrete Math.* **51**, 31–38 (2016)
- Mastroeni, L., Naldi, M.: Compensation policies and risk in service level agreements: a Value-at-Risk approach under the ON-OFF service model. In: *Proceedings of 7th International ICQT Workshop on Advanced Internet Charging and QoS Technology ICQT 2011*, Paris, France (2011)
- Ackermann, T.: *IT Security Risk Management. Perceived IT Security Risks in the Context of Cloud Computing*. Springer Fachmedien, Wiesbaden (2013)
- Arratia, A.: *Computational Finance. An Introductory Course with R*. Atlantis Studies in Computational Finance and Financial Engineering. Atlantis Press, Paris (2014)
- Franken, U.: Optimal IT service availability: shorter outages, or fewer? *IEEE Trans. Netw. Serv. Manag.* **9**(1), 22–33 (2012)
- Vasseur, J.P., Pickavet, M., Demeester, P.: *Network Recovery. Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. Morgan Kaufmann, San Francisco (2004)
- Schupke, D.A.: Multilayer and multidomain resilience in optical networks. *Proc. IEEE* **100**(5), 1140–1148 (2012)
- Elneagaard, N.K., Stordahl, K.: Modelling uncertainty and risk in telecommunication investment project. *Elektronikk* **104**(3/4), 119–135 (2008)
- Wheeler, E.: *Security Risk Management*. Syngress, Waltham (2011)
- Teixeira, A., Sou, K.C., Sandberg, H., Johansson, K.H.: Secure control systems. A quantitative risk management approach. *IEEE Control Syst. Mag.* **35**(1), 24–45 (2015)
- Shin, J., Son, H., Khalil ur, R., Heo, G.: Development of a cyber security risk model using Bayesian networks. *Reliab. Eng. Syst. Saf.* **134**, 208–217 (2015)
- Jing, Y., Ahn, G.J., Zhao, Z., Hu, H.: Towards automated risk assessment and mitigation of mobile applications. *IEEE Trans. Dependable Sec. Comput.* **12**(5), 571–584 (2015)
- Dabbebi, O., Badonnel, R., Festor, O.: An online risk management strategy for VoIP enterprise infrastructures. *J. Netw. Syst. Manag.* **23**(1), 137–162 (2015)
- Wang, J., Chaudhury, A., Rao, H.R.: A Value-at-Risk approach to information security investment. *Inf. Syst. Res.* **19**(1), 106–120 (2008)

21. Cao, Z., Guan, Z., Chen, Z., Hu, J.B., Tang, L.Y.: Towards risk evaluation of Denial-of-Service vulnerabilities in security protocols. *J. Comput. Sci. Technol.* **25**(2), 375–387 (2010)
22. Mello, D.A.A., Schupke, D.A., Waldman, H.: A matrix-based analytical approach to connection unavailability estimation in shared backup path protection. *IEEE Commun. Lett.* **9**(9), 844–846 (2005)
23. Heegaard, P.E., Trivedi, K.S.: Network survivability modeling. *Comput. Netw.* **53**(8), 1215–1234 (2009)
24. Distefano, S., Trivedi, K.S.: Non-Markovian state-space models in dependability evaluation. *Qual. Reliab. Eng. Int.* **26**(2), 225–239 (2013)
25. Ghosh, R., Kim, D., Trivedi, K.S.: System resiliency quantification using non-state-space and state-space analytic models. *Reliab. Eng. Syst. Saf.* **116**, 109–125 (2013)
26. Dikbiyik, F., Tornatore, M., Mukherjee, B.: Minimizing the risk from disaster failures in optical backbone networks. *J. Lightwave Technol.* **32**(18), 3175–3183 (2014)
27. Kuusela, P., Norros, I.: Dynamic approach to Service Level Agreement risk. In: *Proceedings of 9th International Conference on Design of Reliable Communication Networks DRCN 2013*, Budapest, Hungary (2013)
28. González, A.J., Helvik, B.E., Tiwari, P., Becker, D.M., Wittner, O.J.: GEARSHIFT: Guaranteeing availability requirements in SLAs using hybrid fault tolerance. In: *Proceedings of 2015 IEEE Conference on Computer Communications INFOCOM 2015*, Hong Kong, China (2015)
29. Cholda, P., Følstad, E.L., Helvik, B.E., Kuusela, P., Naldi, M., Norros, I.: Towards risk-aware communications networking. *Reliab. Eng. Syst. Saf.* **109**, 160–174 (2013)
30. Markopoulou, A., Iannaccone, G., Bhattacharyya, S., Chuah, C.N., Ganjali, Y., Diot, C.: Characterization of failures in an operational IP backbone network. *IEEE/ACM Trans. Netw.* **16**(4), 749–762 (2008)
31. Kuusela, P., Norros, I.: On/off process modeling of IP network failures. In: *Proceedings of 40th Annual IEEE/IFIP International Conference on Dependable Systems and Networks DSN 2010*, Chicago, IL (2010)
32. González, A.J., Helvik, B.E., Hellan, J.K., Kuusela, P.: Analysis of dependencies between failures in the UNINETT IP backbone network. In: *Proceedings of 16th Pacific Rim International Symposium on Dependable Computing PRDC 2010*, Tokyo, Japan (2010)
33. Uchida, M.: Statistical characteristics of serious network failures in Japan. *Reliab. Eng. Syst. Saf.* **131**, 126–134 (2014)
34. Garraghan, P., Moreno, I.S., Townend, P., Xu, J.: An analysis of failure-related energy waste in a large-scale cloud environment. *IEEE Trans. Emerg. Top. Comput.* **2**(2), 166–180 (2014)
35. Takács, L.: On certain sojourn time problems in the theory of stochastic processes. *Acta Math. Acad. Sci. Hung.* **1–2**(8), 43–48 (1957)
36. Kijima, M.: *Markov Processes for Stochastic Modeling*. Stochastic Modeling Series. Springer, New York (1997)
37. Hedwig, M., Malkowski, S., Neumann, D.: Risk-aware Service Level Agreement design for enterprise information systems. In: *Proceedings of 45th Hawaii International Conference on System Sciences HICSS-45*, Grand Wailea, Maui, HI (2012)
38. Mastroeni, L., Naldi, M.: Violation of service availability targets in Service Level Agreements. In: *Proceedings of Federated Conference on Computer Science and Information Systems FedCSIS 2011*, Szczecin, Poland (2011)
39. Xia, M., Tornatore, M., Martel, C.U., Mukherjee, B.: Risk-aware provisioning for optical WDM mesh networks. *IEEE/ACM Trans. Netw.* **19**(3), 921–931 (2011)
40. Trivedi, K.S.: *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. Wiley, New York (2001)
41. Dikbiyik, F., Reaz, A.S., De Leenheer, M., Mukherjee, B.: Minimizing the disaster risk in optical telecom networks. In: *Proceedings of Optical Fiber Communication and the National Fiber Optic Engineers Conference OFC/NFOEC 2012*, Los Angeles, CA (2012)
42. Snedaker, S., Rima, C.: *Business Continuity and Disaster Recovery Planning for IT Professionals*. Syngress, Waltham (2014)
43. Clemente, R., Bartoli, M., Bossi, M.C., D'Orazio, G., Cosmo, G.: Risk management in availability SLA. In: *Proceedings of 5th International Workshop on the Design of Reliable Communication Networks DRCN 2005*, Lacco Ameno, Island of Ischia, Italy (2005)
44. Olson, D.L., Wu, D.: The impact of distribution on Value-at-Risk measures. *Math. Comput. Model.* **58**(9–10), 1670–1676 (2013)



45. Ahmed, M.S., Al-Shaer, E., Taibah, M., Khan, L.: Objective risk evaluation for automated security management. *J. Netw. Syst. Manag.* **19**(3), 343–366 (2011)
46. Tsai, H.Y., Huang, Y.L.: An analytic hierarchy process-based risk assessment method for wireless networks. *IEEE Trans. Reliab.* **60**(4), 801–816 (2011)
47. Alexander, C., Sarabia, J.M.: Quantile uncertainty and Value-at-Risk model risk. *Risk Anal.* **32**(8), 1293–1308 (2012)
48. MacKenzie, C.A.: Summarizing risk using risk measures and risk indices. *Risk Anal.* **44**(12), 2143–2162 (2014)
49. Mastroeni, L., Naldi, M.: Options and overbooking strategy in the management of wireless spectrum. *Telecommun. Syst.* **48**(1–2), 31–42 (2011)
50. González, A.J., Helvik, B.E.: SLA success probability assessment in networks with correlated failures. *Comput. Commun.* **36**(6), 708–717 (2013)
51. Sun, L., Hong, L.J.: A general framework of importance sampling for Value-at-Risk and Conditional Value-at-Risk. In: *Proceedings of 2009 Winter Simulation Conference WSC 2009*, Austin, TX (2009)
52. Göb, R.: Estimating Value at Risk and Conditional Value at Risk for count variables. *Qual. Reliab. Eng. Int.* **27**(5), 659–672 (2011)
53. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. *Math. Finance* **9**(3), 203–228 (1999)
54. Alexander, C. (ed.): *Value-at-Risk models*. In: *Market Risk Analysis*, vol. IV. Wiley, New York (2008)
55. Mansini, R., Ogryczak, W., Speranza, M.G.: Twenty years of linear programming based portfolio optimization. *Eur. J. Oper. Res.* **234**(2), 518–535 (2014)
56. Alexander, G.J., Baptista, A.M., Yan, S.: Bank regulation and stability: an examination of the Basel market risk framework. In: *Proceedings of Joint Fall Conference Basel III and Beyond: Regulating and Supervising Banks in the Post-Crisis Era*, Eltville am Rhein, Germany (2011)
57. Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and quantile risk measures. *Int. Trans. Oper. Res.* **9**(5), 661–680 (2002)
58. Balakrishnan, N., Limnios, N., Papadopoulos, C.: Basic probabilistic models in reliability. In: Balakrishnan, N., Rao, C.R. (eds.) *Advances in Reliability, Handbook of Statistics*, vol. 20, chap. 1, pp. 1–42. Elsevier, Oxford (2001)
59. Dayar, T.: *Analyzing Markov Chains Using Kronecker Products. Theory and Applications*. Springer Briefs in Mathematics. Springer, New York (2013)
60. Orłowski, S., Wessälly, R., Pióro, M., Tomaszewski, A.: SNDlib 1.0—Survivable Network Design Library. *Networks* **55**(3), 276–286 (2010)
61. Chen, S.X., Hall, P.: Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Stat.* **21**(3), 1166–1181 (1993)

**Krzysztof Rusek** is a PhD candidate at the Department of Telecommunications, AGH University of Science and Technology. He received M.Sc. in Electronics and Telecommunications in 2009 from the AGH University of Science and Technology. His main interests are performance evaluation of communications systems, queuing theory, computer vision, and machine learning.

**Piotr Guzik** is a PhD candidate at the Department of Telecommunications of AGH University of Science and Technology. He received his M.Sc. degree in Astronomy from Jagiellonian University in 2009 (with honors). He has also received M.Sc. degree in Applied Computer Science from AGH University of Science and Technology. His research interests include machine learning, image processing, digital watermarking, and computer vision.

**Piotr Cholda** obtained a doctorate in Telecommunications in 2006 from AGH University of Science and Technology. Then, he joined the Department of Telecommunications there, and now he works as an Assistant Professor. He specializes in design of communications networks, recently he has focused on risk-based networking.